

HR Analysis & Performance Insights

Project Overview

This project focuses on analyzing HR data to understand employee performance, attrition, and workforce distribution. The main goal is to identify why employees leave the company and which departments, job roles, or locations are most affected.

A Power BI dashboard was created to show key HR metrics such as total employees, attrition rate, employee exits by department, performance level, and job role. The dashboard helps HR teams quickly understand trends and make better decisions to improve employee retention and performance.

Dataset Description

This dataset contains HR employee information such as department, job role, gender, performance score, engagement score, start date, exit date, and attrition status.

The data was cleaned and analyzed using Python, SQL, and Power BI to build an interactive HR dashboard.

Exploratory Data Analysis using Python

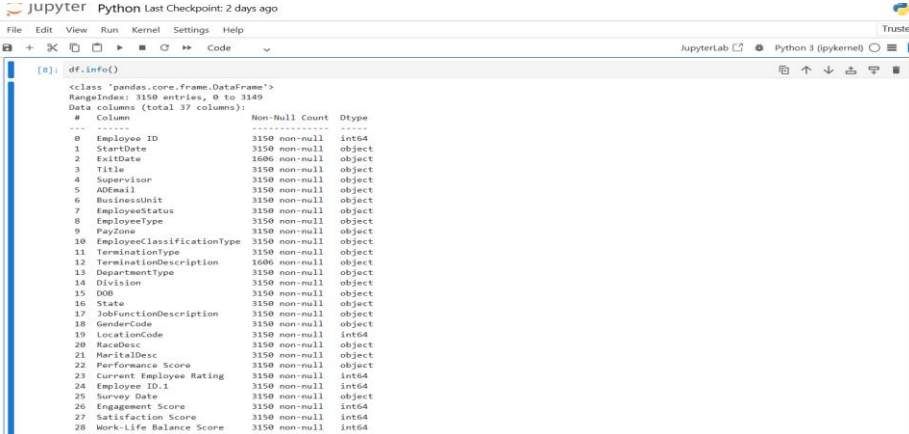
Data preparation and cleaning were performed using Python:

1. Data Loading

Loaded the HR dataset into Python using pandas.

2. Data Understanding

Checked dataset structure, column names, and data types using `df.info()` and `df.head()`.

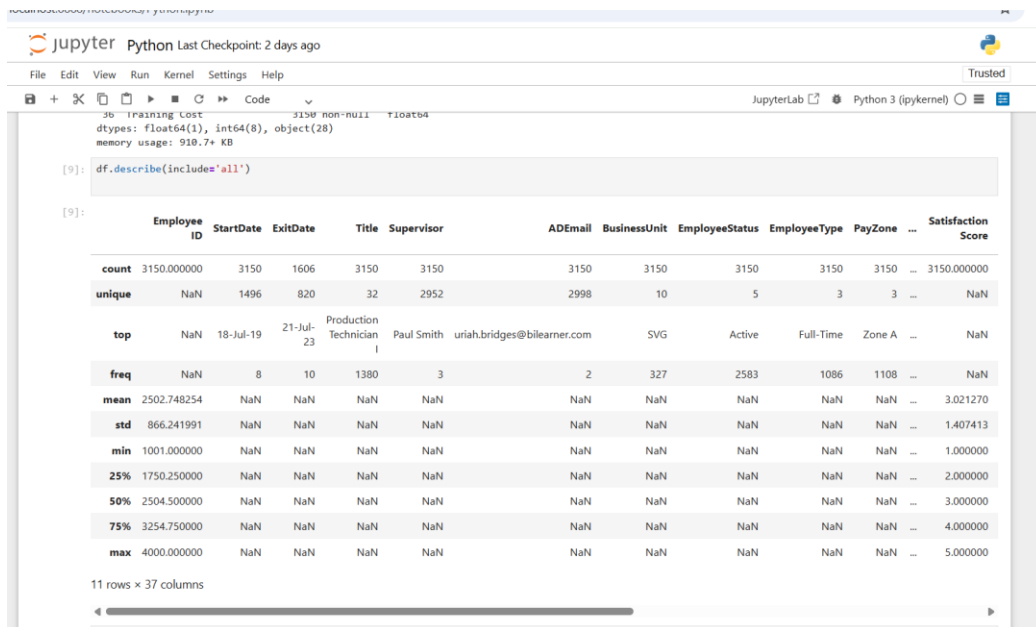


```
[8]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3150 entries, 0 to 3149
Data columns (total 37 columns):
 #   Column                                  Non-Null Count  Dtype  
---  --   ---                                  ---
 0   Employee ID                            3150 non-null   int64  
 1   StartDate                             3150 non-null   object  
 2   ExitDate                              1606 non-null   object  
 3   Title                                  3150 non-null   object  
 4   Supervisor                            3150 non-null   object  
 5   ADEmail                                3150 non-null   object  
 6   BusinessUnit                           3150 non-null   object  
 7   EmployeeStatus                         3150 non-null   object  
 8   EmployeeType                           3150 non-null   object  
 9   PayZone                                3150 non-null   object  
10  EmployeeClassificationType             3150 non-null   object  
11  TerminationType                        3150 non-null   object  
12  TerminationDescription                  1606 non-null   object  
13  DepartmentType                         3150 non-null   object  
14  Division                               3150 non-null   object  
15  DOB                                     3150 non-null   object  
16  State                                   3150 non-null   object  
17  JobFunctionDescription                  3150 non-null   object  
18  GenderCode                             3150 non-null   object  
19  LocationCode                           3150 non-null   int64  
20  RaceDesc                               3150 non-null   object  
21  MaritalDesc                            3150 non-null   object  
22  Performance Score                      3150 non-null   object  
23  Current Employee Rating                 3150 non-null   int64  
24  Employee ID.1                           3150 non-null   int64  
25  Survey Date                            3150 non-null   int64  
26  Engagement Score                       3150 non-null   int64  
27  Satisfaction Score                     3150 non-null   int64  
28  Work-Life Balance Score                 3150 non-null   int64  
29  Termination Reason                      3150 non-null   object
```

3. Summary Statistics

Used `df.describe()` to understand distributions of numerical columns like age, scores, and ratings.



The screenshot shows a JupyterLab interface with a Python kernel. The code cell contains `df.describe(include='all')`. The output is a summary of the data frame, including counts, unique values, top values, frequencies, and statistical measures like mean, standard deviation, minimum, percentiles, and maximum for various columns.

	Employee ID	StartDate	EndDate	Title	Supervisor	ADEmail	BusinessUnit	EmployeeStatus	EmployeeType	PayZone	...	Satisfaction Score
count	3150.000000	3150	1606	3150	3150	3150	3150	3150	3150	3150	...	3150.000000
unique	NaN	1496	820	32	2952	2998	10	5	3	3	...	NaN
top	NaN	18-Jul-19	21-Jul-23	Production Technician I	Paul Smith	urliah.bridges@bilearner.com	SVG	Active	Full-Time	Zone A	...	NaN
freq	NaN	8	10	1380	3	2	327	2583	1086	1108	...	NaN
mean	2502.748254	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	3.021270
std	866.241991	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	1.407413
min	1001.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	1.000000
25%	1750.250000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	2.000000
50%	2504.500000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	3.000000
75%	3254.750000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	4.000000
max	4000.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	5.000000

11 rows x 37 columns

4. Missing Value Check

Identified missing values and handled them appropriately to maintain data quality.

5. Data Cleaning

Renamed columns to a standard format and fixed inconsistent values.

6. Feature Engineering

Created new columns such as age, age_group, tenure_group, attrition, and performance_level.

7. Data Validation

Verified the correctness of newly created columns.

8. Data Export

Loaded the cleaned dataset into PostgreSQL for SQL analysis and Power BI reporting.

Data Analysis using SQL

We performed structured analysis in PostgreSQL to answer key HR business questions:

1. **Total Employees Who Left** – Counted the total number of employees who have exited the company.

2. **Department-wise Attrition** – Identified departments with the highest attrition rate.
3. **Gender-based Attrition** – Compared whether men or women are leaving the company more frequently.
4. **Location-wise Attrition** – Found offices or locations with the highest employee exits.
5. **Average Tenure Before Exit** – Calculated the average time employees stayed before leaving the company.
6. **New Joinees in Last Year/Quarter/Month** – Counted how many employees joined in recent periods.
7. **Performance Level vs. Attrition** – Checked if low-performing employees leave more than high-performing ones.
8. **Current Employees by Department** – Counted the number of active employees in each department.
9. **Job Roles with High Attrition** – Identified job roles or titles that experience the highest turnover.
10. **Supervisors with High Attrition Teams** – Found supervisors whose teams have the highest attrition rate.

HR Analytics Dashboard – Power BI

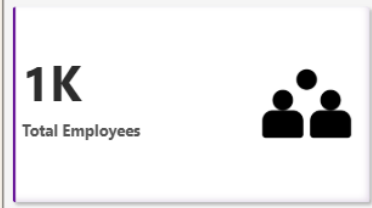
Overview

Created an interactive HR dashboard in Power BI to help HR teams monitor workforce trends and make data-driven decisions.

Key Highlights

- **Employee Overview:** Shows total employees, number of exits, attrition rate, and average tenure.
- **Attrition Insights:**
 - Departments with highest exits
 - Male vs Female attrition
 - Performance level impact on attrition
- **Workforce Details:**
 - High-risk job roles
 - Locations with more employee exits

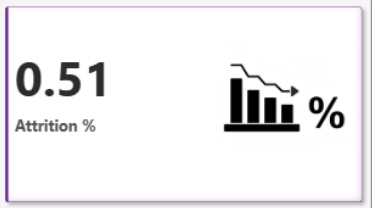
HR Analysis & Performance Dashboard

Total Employees

Attrition Count



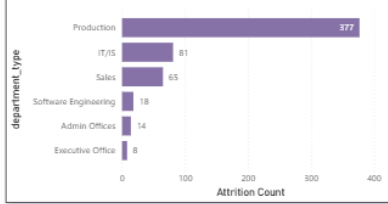
Attrition %



Attrition Count by department_type

A horizontal bar chart titled 'Attrition Count by department_type'. The y-axis is labeled 'department_type' and lists six categories: Production, IT/IS, Sales, Software Engineering, Admin Offices, and Executive Office. The x-axis is labeled 'Attrition Count' and ranges from 0 to 400 with major grid lines every 100 units. The bars are purple, and the exact count for each bar is displayed at its end.

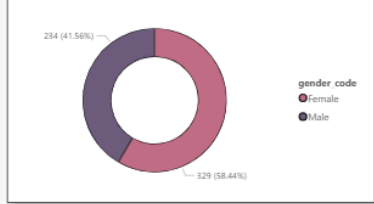
department_type	Attrition Count
Production	377
IT/IS	81
Sales	65
Software Engineering	18
Admin Offices	14
Executive Office	8



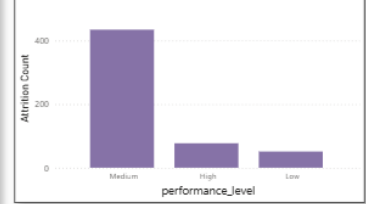
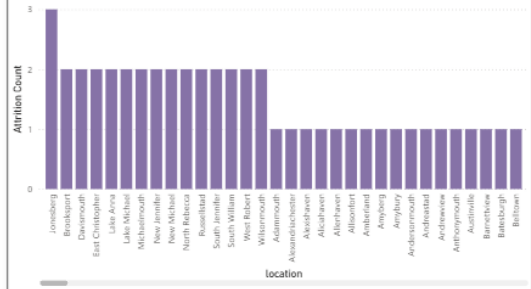
Attrition Count by gender_code

A donut chart titled "Attrition Count by gender_code". The chart is divided into two segments: a larger pink segment representing females and a smaller purple segment representing males. The female segment is labeled with "329 (58.44%)" and the male segment is labeled with "234 (41.56%)". To the right of the chart is a legend titled "gender_code" with a pink circle next to "Female" and a purple circle next to "Male".

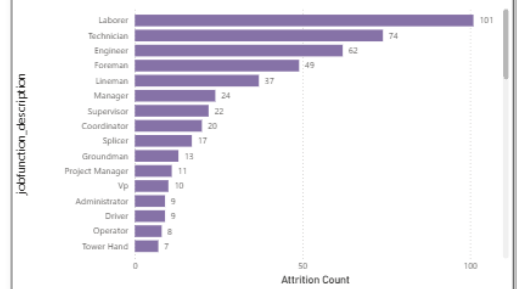
gender_code	Count	Percentage
Female	329	58.44%
Male	234	41.56%



performance_level	Attrition Count
Medium	440
High	80
Low	50

[illegible]

jobfunction_description	Attrition Count
Laborer	101
Technician	74
Engineer	62
Foreman	49
Lineman	37
Manager	24
Supervisor	22
Coordinator	20
Splicer	17
Groundman	13
Project Manager	11
Vp	10
Administrator	9
Driver	9
Operator	8
Tower Hand	7



Business Recommendations

- Improve retention strategies in high-attrition departments
- Focus on employee engagement for low-performing groups
- Strengthen retention programs for early-tenure employees
- Monitor location-wise attrition trends regularly

Conclusion

The HR Analysis & Performance Dashboard provides a clear view of employee attrition, performance trends, and workforce distribution. Using Python, SQL, and Power BI, key HR metrics were analyzed and visualized to identify high-risk departments, job roles, and locations. This dashboard enables HR teams to make data-driven decisions for improving employee retention, performance management, and overall workforce planning.